

# A Review on Human Motion Detection Techniques for ATM-CCTV Surveillance System

Tasriva Sikandar and Kamarul Hawari Ghazali

**Abstract**— The rising rate of Auto Teller Machine (ATM) robbery cases in Malaysia has drawn our attention. Despite having the CCTV surveillance system, the ATM robbery has been occurring in an alarming rate recently. One of the reason is the CCTV recorded video in ATM is limited to use as forensic evidence after crime occurrence while the video for serving as a real time alarm for the security guard (usually occupied with number of monitors in the control room) during robbery has been disregarded. Real time analysis of human motion from CCTV video data can be used to notify the security guard about a suspicious individual committing burglary and thus gives chance to immediately cease the criminal red-handed. In the process of human motion analysis, at first human must be detected in an image or video following three stages namely *i)* environment/background modeling, *ii)* motion segmentation and *iii)* object classification. In this paper we address existing human detection methods that can be effectively used in CCTV surveillance of ATM. Specifically; first we investigate the recent records of ATM robbery cases in all over the world including Europe, America and Asia. Next, we discuss the methods used in the three steps of human motion detection explicitly. The main purpose of the study is to provide a general overview of human motion detection methods used in ATM-CCTV surveillance system.

**Index Terms**—ATM Robbery, CCTV Surveillance, Real Time Video Analysis, Human Detection.

## I. INTRODUCTION AND BACKGROUND STUDY

ATM robberies have been matter of serious concern in most of the countries of the world for many years[1]–[4]. Now it has become a worrying issue in Malaysia because of the frequently occurred ATM robberies in the last few years. There 7345 and 8421 ATM robbery cases were filed in whole Europe in 2013 and 2014 respectively [3]. Also, 8322 cases were recorded in India in 2012 [4]. Fig. 1 shows data records of ATM robberies in Malaysia in last five years [5], [6]. It is worth to mention that, there is at least one ATM robbery per 3.5 million transactions in USA and Australia every year [1], [2]. Despite knowing the fact, ATMs are under the CCTV surveillance, robbers are not afraid of committing criminal offences. One of the reason could be the recorded video of ordinary CCTV is currently used as forensic evidence not as an “active real time media” to alert the unaware security guard about the ongoing robbery. Using the CCTV video for a real time alarm system has not been considered yet to improve the

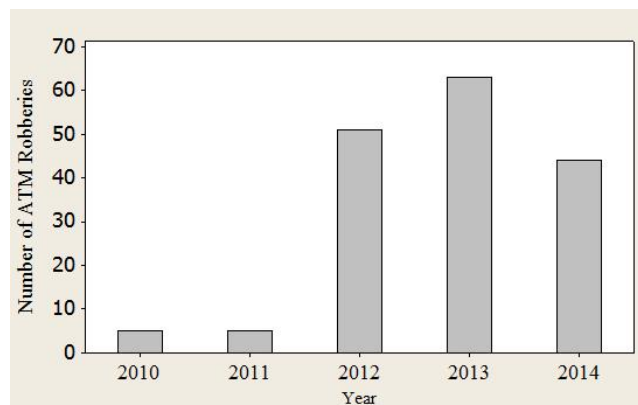


Fig. 1. ATM robbery occurrence in Malaysia (2010-2014)

ATM security. Real time analysis of human motion from CCTV video data can be used for early detection of suspicious person to alert the security guard, usually occupied with so many monitors in the control room, for redrawing their attention to stop the ongoing ATM robbery.

Human Motion analysis has been utilized in several surveillance projects such as Video Motion Detector (VMD) [7] Amethyst [8], Virtual Guard [9], Video Surveillance and Monitoring (VSAM) [10], W<sup>4</sup> [11], and Human Motion Detection Framework [12]. Detection of human in an image or video is the first and vital step in the process of human motion analysis. Different human motion detection methods have been surveyed in [13]–[17] as part of the human motion analysis process framework in addition to [18], [19] where surveillance issue was focused. The process of human detection in surveillance video usually involves environment modeling, motion segmentation and object classification. These stages may overlap with each other during processing. In this paper we study different methods of human detection in video that can be effectively utilized in CCTV surveillance of ATM overcoming the environmental challenging factors. The focus of this paper is on compiling and presenting a general overview of the previous investigations on human motion detection methods which are feasible to use in ATM-CCTV surveillance environment, instead of giving detailed mathematically comprehensive summaries of the individual study.

## II. ENVIRONMENT FACTORS AND BACKGROUND MODELING

Background models of an ATM environment can be classified into 2-D models in image plane and 3-D models in real world coordinates. Due to the simplicity, 2-D models h-

Manuscript received March 31, 2015.

Tasriva Sikandar is with the Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang, 26600-Pekan, Malaysia

Kamarul Hawari Ghazali, Associate Professor and Dean, is with the Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang, 26600-Pekan, Malaysia

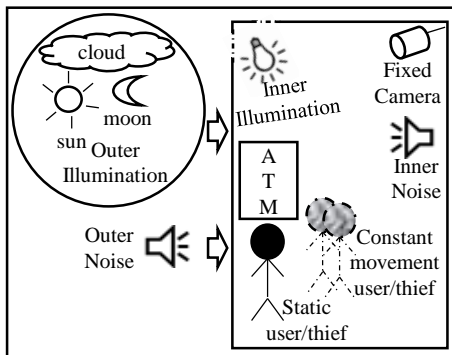


Fig. 2. Factors contributing to ATM environment.

-ave more applications. In addition, good understanding of background/environment factors is essential for active construction and updating of background/environmental modeling in visual surveillance or more specifically in human detection in image or video. Fig. 2 presents factors contributing to ATM environment/background modeling.

The main problem of the still cameras is recovering and updating background images from a dynamic sequence automatically. Unfavorable factors, such as inner and outer illumination variance, shadows, slow movement and constant/repetitive movement of user/thief, thief hidden behind another object/occlusion, camouflage and noise (outer, inner, video) in the ATM environment bring many difficulties to the acquirement and updating of background images. Many algorithms are available to overcome these problems [20]–[27]. A reasonably effective solution was given in [20] to update the background image with only those pixels not identified as moving objects. But the approach cannot handle shadow and slow movement.

In [21], a probabilistic approach is applied for each pixel classification, than each pixel appearance is modeled as a mixture of Gaussian and is learned using incremental Expectation Maximization (EM) algorithm. This approach can handle shadow, but is not suitable for extreme level of lighting condition, is subject to noise effect and is susceptible to small motion effects.

The method described in [22], [23] recover and update background images simply by modeling the values of a particular pixel as a mixture of Gaussians, rather than explicitly modeling the values of all the pixels as one particular type of distribution. This approach can deal with lighting changes, repetitive motions and long-term scene changes.

[24] demonstrated a technique which models each pixel value with a Kalman filter to neutralize the effect of illumination variance in updating background images. It is also useful for smaller image resolution which speed up the system. In [25], background maintenance is carried out at three levels: the pixel level, the region level, and the frame level describe as in Fig. 3. It can manage illumination variance, camouflage and constant motion.

In [26], background modeling is managed by, for each pixel, taking a set of values in the past at the same location or in the neighborhood, then comparing this set to the current pixel value in order to determine whether that pixel belongs to the background, and adapts the model by choosing randomly

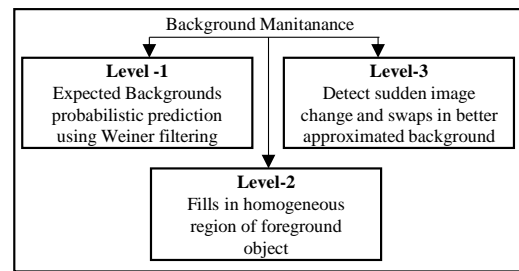


Fig. 3. Background Maintenance approach in [25]

which values to substitute from the background model. The approach is noise resilience. [27] is a thesis presenting a template matching method robust against image noise. Studies show that, a statistical model [11] and an adaptive background model with color and gradient information [28] can reduce influence of shadows and unreliable color cues.

### III. MOTION SEGMENTATION

In motion segmentation, a video or image is decomposed in moving object region and background. For better understanding we outlined the algorithms under some categories arranged in accordance to their fundamental concept of the approach.

#### A. Background Subtraction

Background subtraction detects moving object region in an image by differencing between current image and a reference background image in a pixel-by-pixel manner. Although background subtraction methods [22], [28]–[35] have become popular for motion segmentation for simplicity, these techniques suffer from shortcomings with respect to the challenges of video surveillance. All the above studies have different type of background model and procedure to update the background model. The method described in [33] is a spatio-temporal technique and can cope with inconsistencies due to lighting changes. [34], [35] presented an adaptive background model based on Kalman filtering. They can adjust to the temporal changes of lighting and can handle shadow problem. In [36], performance of nine background subtraction methods were compared. The comparison was carried out in accordance to their ability to fulfill the visual surveillance challenges.

#### B. Statistical Methods

To construct more improved and dynamic background models, statistical approaches can be applied. Those techniques have the advantages of updating the background by simply updating the background statistics during processing [15]. For example, each pixel in the current image can be classified into foreground or background by comparing the statistics of the current background model. Statistical methods are extensively used in motion segmentation of dynamic scene due to its robustness to noise, shadow, and change of lighting conditions, etc. Statistical methods in [11], [22], [23], [28], [29] are some examples to be mentioned.

#### C. Temporal Differencing

In the temporal differencing approach, pixel-wise difference between two or three consecutive frames in an image sequence

are used to detect moving regions. Although temporal differencing methods are very swift in computation and adaptive to dynamic environments, generally does a poor job of extracting the entire relevant feature pixels. Possibly this approach generates holes inside moving entities. As an example of this approach the algorithm described in [37] is capable of detection motion despite occlusion and the algorithm of [38] is robust in presence of noise.

#### D. Optical Flow

The characteristics of flow vectors of moving objects over time has been used to extract moving regions in an image sequence in the optical flow based motion segmentation methods. These methods are particularly used for gait analysis and are able to detect independently moving objects. However, most of the methods require highly complex computation and fails to detect stable objects. In addition, most of the methods are sensitive to noise and occlusion. Such examples of optical flow approaches are [39], [40]. Motion segmentation is an extensively large area of study. Besides the above mentioned methods, there are apparently some other approaches which can be used in ATM CCTV surveillance.

### IV. OBJECT CLASSIFICATION

Generally, object classification may not be essential step under some visual surveillance system where the moving objects are known to be human. Although ATM CCTV surveillance system seems to match with this kind of situation, it is very obvious to encounter pedestrian and animals (cat, dog) roaming around the area of the ATM. In addition, accurate classifying of moving object is essential for further analyzation of human behavior. Classifying the moving object after segmentation of moving regions for ATM CCTV video is an essential task. For the convenience of discussion object classification methods are outlined under two groups.

#### A. Classification based on shape

In shape based classification methods, moving objects are classified using simple shape primitives such as point, box, silhouette and blob of motion region. [10], [37] makes use of blob features. The study in [10] classifies moving-object blobs into four classes: single human, vehicles, human groups, and clutter classification is using a viewpoint-specific three-layer neural network classifier. He took image blob disperseness, image blob area, apparent aspect ratio of the blob bounding box, etc as parameter for classification. In addition, [37] proposed a technique where the image dispersedness and area of image blobs were taken as classification metrics to classify all moving-object blobs into humans, vehicles and clutter. [41] used simple shape parameters of human silhouette patterns to separate humans from other moving objects. On the other hand, [42] proposed a method starting with silhouettes and detected the body parts using a method inspired by the W4-system [11].

#### B. Classification based on motion

In general, non-rigid articulated human motion shows a periodic property [15]. Motion based classification methods makes use of this periodic motion property to classify moving objects in an image. [43] proposed a self-similarity based

technique that can robustly detect and analyze periodic motion. By tracking an interesting moving object, its self-similarity is computed as it evolves over time. In addition, time-frequency analysis can also be applied to detect and characterize the periodic motion for tracking and classification of moving objects. The technique is suitable for both fix and moving camera. Although optical flow has been considered in applicable in real time visual surveillance for their highly complex computation, Lipton's work [44] describes a reliable and less computation time demanding flow-based technique called "dynamic region matching". It uses residual flow to analyze rigidity and periodicity of moving objects in real time. It is expected that rigid objects present little residual flow, whereas a non-rigid moving object such as a human being has a higher average residual flow and even display a periodic component. Based on this useful cue, human motion is distinguished from motion of other objects, such as vehicles.

### V. CONCLUSION

In this paper we address image processing techniques in the ATM-CCTV video for analyzing the human motion to give alert to the security guard. As the first and vital step, in this paper, we attempted to survey human detection methods useful for ATM-CCTV surveillance. We analyzed ATM environment factors that influence human detection process and presented a general overview of different human detection techniques. Overall, after reviewing all existing methods it is apparent that, for successful detection of human in the ATM environment the applied algorithms need to be robust against unfavorable factors such as, inner and outer illumination variance, shadows, slow movement and constant and/or repetitive movement of user, thief hidden behind another object, camouflage and noise (outer, inner, video) in the ATM environment. However, it is hard to find a single method that comes up with all the boons. A combination of several algorithms may work better than any existing method.

### REFERENCES

- [1] M. S. Scott, "Robbery at Automated Teller Machines," Community Oriented Policing Services, U.S. Department of justice, 2002.
- [2] N. Diana and R. G. Smith, "ATM related Robberies," Canberra: Australian Institute of Criminology, 2002.
- [3] "European ATM Fraud Incidents up 15%, driven by low tech crime," European ATM Crime Report. European ATM Security Team (EAST), 2015.
- [4] S. Krishnan, "Invading privacy: Cyber crimes on the rise," PwC India, 2013.
- [5] M. Datuk bin Ibrahim, "Deputy Governor's Speech at the Launch of MEPS ATM Initiative," *Bank Negara Malaysia(Central Bank of Malaysia) speeches and Interviews*, 2013. .
- [6] M. Datuk bin Ibrahim, "Deputy Governor's Keynote Address at the Payment System Forum and Exhibition 2014," *Bank Negara Malaysia(Central Bank of Malaysia) speeches and Interviews*, 2014. .
- [7] S. Aviel, "Video Motion Detector."
- [8] M. Horner, G. Leach, and T. O. Dwyer, "AMETHYST Automatic Alarm Assessment: Operational Experience," in *Security Technology, 2000. Proceedings.IEEE 34th Annual 2000 International Carnahan Conference on. IEEE.*, 2000, pp. 107–112.  
<http://dx.doi.org/10.1109/ccst.2000.891175>
- [9] G. Thiel, "Automatic CCTV surveillance - towards the VIRTUAL GUARD," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 15, no. 7, pp. 3–9, 2000.  
<http://dx.doi.org/10.1109/62.854018>

- [10] R. T. Collins, A. J. Lipton, and T. Kanade, *A System for Video Surveillance and Monitoring*, vol. 2, 2000.
- [11] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Real-time surveillance of people and their activities," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 809–830, 2000.  
<http://dx.doi.org/10.1109/34.868683>
- [12] S. B. Ramli, K. H. B. Ghazali, M. Faiz, B. Mohd, and Z. H. B. Hisahuddin, "Human Motion Detection Framework," pp. 158–161, 2011.
- [13] C. Cdras and M. Shah, "Motion-based recognition: a survey," *Image Vis. Comput.* 13.2, pp. 129–155, 1995.  
[http://dx.doi.org/10.1016/0262-8856\(95\)93154-K](http://dx.doi.org/10.1016/0262-8856(95)93154-K)
- [14] J. K. K. Aggarwal and Q. Cai, "Human motion analysis: a review," *IEEE Nonrigid Articul. Motion Work.*, vol. 73, no. 3, pp. 90–102, 1999.  
<http://dx.doi.org/10.1006/cviu.1998.0744>
- [15] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognit.*, vol. 36, no. 3, pp. 585–601, 2003.  
[http://dx.doi.org/10.1016/S0031-3203\(02\)00100-0](http://dx.doi.org/10.1016/S0031-3203(02)00100-0)
- [16] H. Buxton, "Learning and understanding dynamic scene activity: A review," *Image Vis. Comput.*, vol. 21, no. 1, pp. 125–136, 2003.  
[http://dx.doi.org/10.1016/S0262-8856\(02\)00127-0](http://dx.doi.org/10.1016/S0262-8856(02)00127-0)
- [17] T. B. Moeslund, A. Hilton, and V. Kruger, "A survey of advances in vision-based human motion capture and analysis," *Comput. Vis. Image Underst.*, vol. 104, no. 2–3 SPEC. ISS., pp. 90–126, 2006.
- [18] W. Hu, T. Teiniu, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *Syst. Man, Cybern. Part C Appl. Rev. IEEE Trans.*, vol. 34.3, pp. 334–352, 2004.
- [19] T. Ko, "A Survey on Behaviour Analysis in Video Surveillance Applications," *Video Surveill.*, pp. 279–294, 2011.
- [20] D. Koller, J. Weber, T. Huang, J. Malik, G. Ogasawara, B. Rao, and S. Russell, "Towards robust automatic traffic scene analysis in real-time," *Proc. 12th Int. Conf. Pattern Recognit.*, vol. 1, pp. 126–131, 1994.
- [21] N. Friedman and S. Russell, "Image segmentation in video sequences: a probabilistic approach," *Proc. Thirteenth. Conf. Uncertain. Artif. Intell.*, pp. 175–181, 1997.
- [22] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," *Proceedings. 1999 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (Cat. No PR00149)*, vol. 2, pp. 246–252, 1999.  
<http://dx.doi.org/10.1109/cvpr.1999.784637>
- [23] L. Lee, W. E. L. Grimson, C. Stauffer, and R. Roman, "Using adaptive tracking to classify and monitor activities in a site," in *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on. IEEE*, 1998, pp. 22–29.
- [24] C. Ridder, O. Munkelt, and H. Kirchner, "Adaptive background estimation and foreground detection using kalman-filtering," in *Proceedings of International Conference on recent Advances in Mechatronics*, 1995.
- [25] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: principles and practice of background maintenance," *Comput. Vision, 1999. Proc. Seventh IEEE Int. Conf.*, vol. 1, pp. 255–261 vol.1, 1999.  
<http://dx.doi.org/10.1109/ICCV.1999.791228>
- [26] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1709–1724, 2011.  
<http://dx.doi.org/10.1109/TIP.2010.2101613>
- [27] D. Thanh Nguyen, "Human detection in images and videos," University of Wollongong, 2012.
- [28] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, "Tracking Groups of People," *Comput. Vis. Image Underst.*, vol. 80, no. 1, pp. 42–56, 2000.  
<http://dx.doi.org/10.1006/cviu.2000.0870>
- [29] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 780–785, 1997.  
<http://dx.doi.org/10.1109/34.598236>
- [30] S. Arseneau and J. R. Cooperstock, "Real-time image segmentation for action recognition," *1999 IEEE Pacific Rim Conf. Commun. Comput. Signal Process. (PACRIM 1999). Conf. Proc. (Cat. No.99CH36368)*, pp. 86–89, 1999.  
<http://dx.doi.org/10.1109/pacrim.1999.799484>
- [31] H. Z. Sun, T. Feng, and T. N. Tan, "Robust extraction of moving objects from image sequences," in *Proceedings of the Fourth Asian Conference on Computer Vision, Taiwan, 2000*, pp. 961–964., 2000.
- [32] A. Elgammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," *Comput. Vision—ECCV 2000*, vol. 1843, pp. 751–767, 2000.  
[http://dx.doi.org/10.1007/3-540-45053-x\\_48](http://dx.doi.org/10.1007/3-540-45053-x_48)
- [33] Y. H. Yang and M. D. Levine, "The background primal sketch: An approach for tracking moving objects," *Mach. Vis. Appl.*, vol. 5, no. 1, pp. 17–34, 1992.  
<http://dx.doi.org/10.1007/BF01213527>
- [34] K. P. Karmann and A. Brandt, "Moving object recognition using an adaptive background memory," V. Cappellini (Ed.), *Time-Varying Image Process. Mov. Object Recognit.*, vol. 2, no. Elsevier, Amsterdam, The Netherlands, 1990.
- [35] M. Kilger, "A shadow handler in a video-based real-time traffic monitoring system," *1992 Proc. IEEE Work. Appl. Comput. Vis.*, pp. 11–18, 1992.  
<http://dx.doi.org/10.1109/acv.1992.240332>
- [36] S. Brutzer, B. Höferlin, and G. Heidemann, "Evaluation of background subtraction techniques for video surveillance," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1937–1944, 2011.  
<http://dx.doi.org/10.1109/cvpr.2011.5995508>
- [37] a J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving target classification and tracking from real-time video," *Proc. Fourth IEEE Work. Appl. Comput. Vis. WACV98 Cat No98EX201*, vol. 98, no. 2, pp. 8–14, 1998.
- [38] J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg, "A three-frame algorithm for estimating two-component image motion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, pp. 886–896, 1992.  
<http://dx.doi.org/10.1109/34.161348>
- [39] D. Meyer, J. Denzler, and H. Niemann, "Model based extraction of articulated objects in image sequences\nfor gait analysis," *Proc. Int. Conf. Image Process.*, vol. 3, no. Informatik 5, pp. 78–81, 1997.  
<http://dx.doi.org/10.1109/ICIP.1997.631988>
- [40] D. Meyer and H. Niemann, "Gait classification with HMMs for trajectories of body parts extracted by mixture densities," *Br. Mach. Vis. Conf.*, pp. 459–468, 1998.  
<http://dx.doi.org/10.5244/c.12.46>
- [41] Y. Kuno, T. Watanabe, Y. Shimosakoda, and S. Nakagawa, "Automated detection of human for visual surveillance system," *Proc. 13th Int. Conf. Pattern Recognit.*, vol. 3, pp. 865–869, 1996.
- [42] J. W. Davis and S. R. Taylor, "Analysis and recognition of walking movements," in *Proceedings of 16th International Conference on Pattern Recognition*, 2002, p. 10315.  
<http://dx.doi.org/10.1109/icpr.2002.1044702>
- [43] R. Cutler and L. S. Davis, "Robust real-time periodic motion detection, analysis, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 781–796, 2000.  
<http://dx.doi.org/10.1109/34.868681>
- [44] A. J. Lipton, "Local Application of Optic Flow to Analyse Rigid versus Non-Rigid Motion," *Int. Conf. Comput. Vis. Work. Fram. Appl.*, 1999.



Tasriva Sikandar  
Student: Master of Science  
Universiti Malaysia Pahang  
Faculty of Electrical & Electronics Engineering  
Pekan-26600, Pahang, Malaysia  
e-mail: tasrivasikandar@gmail.com  
Phone :+601121998836

Tasriva Sikandar received her BSc. (Engg.) degree in 2010 from Shahjalal University of Science & Technology, Sylhet, Bangladesh in Computer Science and Engineering. She then worked as a lecturer in Department of Computer Science and Engineering in Southern University Bangladesh, Chittagong, Bangladesh. Now she is pursuing her Master of Electrical Engineering. Her research area is image and video processing applications.



Kamarul Hawari Ghazali  
Associate Professor and Dean,  
Faculty of Electrical and Electronics Engineering  
University Malaysia Pahang  
Pekan-26600, Pahang, Malaysia  
e-mail: kamarul@ump.edu.my  
Phone: +60-94246000  
H/p: +60177712224

Dr. Kamarul Hawari Ghazali received his first degree in 1995 from Universiti Teknologi Malaysia in Electrical Engineering. He then continues his

study in Master of Electrical Engineering and was graduated in 2003. He has PhD in Electronics System Engineering specialist in Computer Vision System. He was graduated in PhD from Universiti Kebangsaan Malaysia. He is currently an Associate Professor in the Faculty of Electrical and Electronics Engineering of Universiti Malaysia Pahang. His research area is in signal and image processing applications, computer vision and intelligent system.